APPLICATION FOR UNITED STATES PATENT

for

NOVEL DNA REPAIR ENZYMES, NUCLEIC ACIDS ENCODING DNA REPAIR ENZYMES AND METHODS OF USING THEM

Inventors: **SEIKI KURAMITSU**, a citizen of Japan
residing at

SHIGEYUKI YOKOYAMA, a citizen of Japan
residing at

Assignee: **Riken**
1-7-22 Suehiro-cho, Turumi-ku
Yokohama-shi
Kanagawa 230-0045 Japan

5

Fish & Richardson P.C.
4350 La Jolla Village Drive, Suite 500
San Diego, CA 92122
Tel.: (858) 678-5070
Fax: (858) 678-5099

ATTORNEY DOCKET:
11283-013001

DATE OF DEPOSIT: August 24, 2001

EXPRESS MAIL NO.: EM EL584939500US US

NOVEL DNA REPAIR ENZYMES, NUCLEIC ACIDS ENCODING DNA REPAIR ENZYMES AND METHODS OF USING THEM

CROSS-REFERENCES TO RELATED APPLICATIONS

The present application claims priority under 35 U.S.C. §119 to Japan Patent Application No. 47762/2001, filed February 23, 2001. The aforementioned application is explicitly incorporated herein by reference in its entirety and for all purposes.

TECHNICAL FIELD

The present invention relates to DNA repair enzymes, genes encoding the enzymes, and methods of DNA repair.

BACKGROUND

Genomic DNA in cells in which all the information necessary for the maintenance of life is written is always undergoing damage caused by various exogenous and endogenous factors. As exogenous factors, ultraviolet light, ionizing radiation and environmental chemical substances may be enumerated, for example. As endogenous factors, several types of active oxygen generated from energy metabolism and oxidation stress may be enumerated, for example. Further, mismatches that do not pair correctly with the template can be generated during DNA replication.

When these damaged sites or mismatches are left without repair, bases in the relevant sites will be different from what they are supposed to be, resulting in inaccurate genetic information, i.e., mutations. If a mutation has occurred in a coding region for a protein, the protein may have lower activity (or even no activity) than the corresponding native protein, or the protein may not be produced at all. If a mutation has occurred in a regulatory region, the level of synthesis of the protein under the control of this region can be abnormally increased or decreased. Further, control by other proteins may become ineffective. These changes may cause apoptosis or abnormal growth, e.g., canceration, in relevant cells.

Since damages or mismatches in DNA affect the life of cells per se and may even affect the life of the individuals to which the cells belong, cells have mechanisms to repair DNA damages or mismatches and thereby to maintain genetic information accurately.

These are called DNA repair mechanisms. There are several types of DNA repair mechanisms, including base excision repair, photoreactivation, nucleotide excision repair, mismatch repair and recombination repair. It is expected that elucidation of DNA repair mechanisms would provide findings useful for the study of diseases such as cancer and the study of effects of environmental factors on living organisms. Furthermore, certain types of proteins involved in DNA repair mechanisms are expected to increase the accuracy of PCR that has become an important technique in various fields beyond the field of molecular biology.

Genes of a number of DNA repair enzymes have already been cloned from various organisms, and three-dimensional structural analysis of proteins has been carried out for some of them. However, most of these studies performed to date are genetic studies, and biochemical studies have been performed little. In order to elucidate DNA repair mechanisms and obtain findings useful in various fields such as medicine, it is necessary to clone all genes involved in DNA repair and to carry out three-dimensional structural analysis and detailed functional analysis of the encoded proteins.

SUMMARY

The invention provides novel DNA repair enzymes, genes encoding the enzymes and methods of DNA repair. As a result of extensive and intensive research toward the solution of the above problem, the present inventors have succeeded in isolation of genes encoding DNA repair enzymes from a highly thermophilic bacterium.

The present invention provides an isolated protein selected from the group consisting of the following (a) and (b): (a) a protein comprising the amino acid sequence as shown in SEQ ID NO: 2, 4, 6 or 8; (b) a protein which comprises the amino acid sequence as shown in SEQ ID NO: 2, 4, 6 or 8 having a deletion(s), substitution(s) or addition(s) of one or several amino acids and which has DNA repair enzyme activity.

The present invention provides a DNA repair enzyme encoded by a nucleic acid, wherein the nucleic acid hybridizes under stringent conditions with a nucleic acid comprising all or a part of the nucleotide sequence as set forth in SEQ ID NO: 1, 3, 5 or 7, or from a complementary strand thereto.

In alternative aspects, the present invention provides DNA repair enzymes comprising an amino acid sequence which has at least 60%, at least 65%, at least 70%, at

least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, homology to the amino acid sequence as shown in SEQ ID NO: 2, 4, 6 or 8 and which has DNA repair enzyme. In one aspect, a BLAST algorithm is used to determine the sequence identities, as described, below.

The present invention provides an isolated gene encoding a DNA repair enzyme comprising a DNA encoding the following protein (a) or (b): (a) a protein comprising the amino acid sequence as shown in SEQ ID NO: 2, 4, 6 or 8; (b) a protein which comprises the amino acid sequence as shown in SEQ ID NO: 2, 4, 6 or 8 having a deletion(s), substitution(s) or addition(s) of one or several amino acids and which has DNA repair enzyme activity.

The present invention provides an isolated gene for a DNA repair enzyme comprising the following DNA (c), (d), (e) or (f): (c) a DNA comprising the nucleotide sequence as shown in SEQ ID NO: 1, 3, 5 or 7; (d) a complementary strand to (a); (e) a DNA which hybridizes under stringent conditions either with a DNA consisting of the nucleotide sequence as shown in SEQ ID NO: 1, 3, 5 or 7 or with a complementary strand thereto, and which encodes or is complementary to a DNA which encodes a protein having DNA repair enzyme activity; (f) a DNA which hybridizes under stringent conditions with a probe prepared either from a DNA consisting of the whole or a part of the nucleotide sequence as shown in SEQ ID NO: 1, 3, 5 or 7 or from a complementary strand thereto, and which encodes or is complementary to a DNA which encodes a protein having DNA repair enzyme activity.

The present invention provides recombinant vector comprising the above-described gene. The recombinant vector can be a plasmid, a recombinant virus, a cosmid, an artificial chromosome, and the like.

The present invention provides a cell transformant comprising the above-described recombinant vector. The cell can be a bacterial cell, an insect cell, a plant cell, a mammalian cell, a yeast cell, and the like. The invention also provides a transgenic non-human animal comprising a nucleic acid or a polypeptide of the invention.

The present invention provides method of producing a DNA repair enzyme, comprising culturing the above-described transformant and recovering the DNA repair enzyme from the resultant culture.

The present invention provides a method of repairing DNA sequence errors, comprising carrying out a DNA synthesis reaction in the presence of the above-described protein. The method can be carried out *in vitro* or *in vivo*.

The present invention provides a method of preventing erroneous synthesis of DNA sequences, comprising carrying out a DNA synthesis reaction in the presence of the above-described protein.

The present invention provides a repair gene-disrupted (i.e., "knockout") strain obtained by transferring into a host a construct comprising a nucleic acid of the invention; in one aspect, a modified gene has been incorporated into the construct. A marker gene may be given with the modified gene, or, in the same construct as the modified gene. As a specific example of a host is a thermophilic bacterium. In one aspect the thermophilic bacterium is a bacterium of the genus *Thermus*, such as *Thermus thermophilus*.

The proteins of the invention can be stable in a temperatures ranging from about 4°C to about 100°C. In one aspect, the proteins of the invention are stable up to 98°C, up to 95°C, up to 90°C, up to 80°C, up to 75°C.

The invention also provides arrays (i.e., a "biochip") comprising a nucleic acid as set forth in SEQ ID NO: 1, 3, 5 or 7, and, arrays comprising a nucleic acid of the invention.

The invention provides a method of screening a composition for its ability to specifically bind to a DNA repair enzyme comprising: (a) contacting the a DNA repair enzyme with the composition, wherein the DNA repair enzyme is a polypeptide encoded by a nucleic acid sequence of the invention; and, (b) determining if the composition specifically binds to the DNA repair enzyme.

The invention provides a method for inhibiting the expression of a DNA repair enzyme encoding nucleic acid in a cell, the method comprising the following steps: (a) providing a nucleic acid operably linked to a promoter that expresses an inhibitory sequence, wherein the inhibitory sequence comprises all or part of a nucleic acid sequence of the invention and is expressed in a form sufficient to inhibit expression of a DNA repair enzyme message; and, (b) expressing the inhibitory nucleic acid in an amount sufficient to inhibit the expression of the DNA repair enzyme encoding nucleic acid in the cell. In one

aspect, the inhibitory sequence comprises an antisense sequence. In one aspect, the inhibitory sequence comprises a ribozyme sequence.

The invention provides a method of expressing a heterologous nucleic acid sequence in a cell comprising: a) transforming the cell with a heterologous nucleic acid operably linked to a promoter, wherein the heterologous nucleic acid comprises a nucleic acid sequence of the invention; and, b) growing the cell under conditions where the heterologous nucleic acid sequence is expressed in the cell.

The invention provides a method for detecting a nucleic acid in a nucleic acid -containing biological sample, the method comprising the following steps: (a) contacting the sample with a nucleic acid probe comprising a nucleic acid sequence of the invention; (b) hybridizing the nucleic acid probe to the nucleic acid in the sample; and, (c) detecting hybridization of the nucleic acids.

The invention provides a fusion protein comprising a first amino acid sequence as set forth in SEQ ID NO: 2, 4, 6 or 8, or a subsequence thereof, and a second heterologous sequence.

The invention provides an isolated antibody specifically reactive with a polypeptide of the invention or a polypeptide encoded by a nucleic acid of the invention. In one aspect, the antibody is a monoclonal antibody. The invention provides a hybridoma cell comprising a monoclonal antibody of the invention.

The details of one or more aspects of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

All publications, GenBank Accession references (sequences), ATCC Deposits, patents and patent applications cited herein are hereby expressly incorporated by reference for all purposes.

DESCRIPTION OF DRAWINGS

Fig. 1 is a diagram showing the function of MutY.

Fig. 2 is a diagram showing the base excision repair mechanism of MutY.

Fig. 3 is a representation of a photograph showing the results of SDS-polyacrylamide gel electrophoresis of MutY.

Fig. 4 is a chart showing the results of gel filtration of MutY.

Fig. 5 is an alignment of the amino acid sequence of the MutY of the invention with amino acid sequences of other MutY proteins.

Fig. 6 is a diagram showing an outline of the method of measurement of MutY activities.

Fig. 7 is a diagram showing the substrate specificity of MutY.

Fig. 8 is a chart showing the absorption spectrum of MutY.

Fig. 9 is a chart showing the CD spectrum of MutY.

Fig. 10 is a chart showing the thermostability of MutY.

Fig. 11 is a diagram showing substrate DNA and 32P-labeled site.

Fig. 12 is a diagram showing the function of RecJ.

Fig. 13 is a representation of a photograph showing the results of SDS-polyacrylamide gel electrophoresis of RecJ.

Fig. 14 is an alignment of the amino acid sequence of the RecJ of the invention with amino acid sequences of other RecJ proteins.

Fig. 15 is a chart showing the CD spectrum of RecJ.

Fig. 16 is a chart showing the thermostability of RecJ.

Fig. 17 is a diagram showing the method of measurement of the exonuclease activity of RecJ.

Fig. 18 shows results of measurement of the exonuclease activity of RecJ.

Fig. 19 shows results of measurement of the exonuclease activity of RecJ (dependency on RecJ concentration).

Fig. 20 shows the effect of etheno-nucleotide upon RecJ activity.

Fig. 21 shows results of measurement of the exonuclease activity of RecJ (fluorescence spectrum).

Fig. 22 shows the results of measurement of the exonuclease activity of RecJ (time course of fluorescence intensity and the degree of fluorescence polarization).

Fig. 23 shows results of measurement of the exonuclease activity of RecJ (dependency on DNA concentration).

Fig. 24 is a diagram showing the reaction pathway of RecF.

Fig. 25 is a representation of a photograph showing the results of SDS-polyacrylamide gel electrophoresis of RecF.

Fig. 26 is a chart showing the results of gel filtration of RecF.

Fig. 27 is an alignment of the amino acid sequence of the RecF of the invention with amino acid sequences of other RecF proteins.

Fig. 28 presents graphs showing the linking of RecF toεDNA.

Fig. 29 is a graph showing ATPase activity.

Fig. 30 is a graph showing ATPase activity (DNA dependency).

Fig. 31 is a diagram showing the nucleotide excision repair mechanism of TRCF.

Fig. 32 is a drawing showing the three-dimensional structure of UvrB.

Fig. 33 is a representation of photographs showing the results of SDS-polyacrylamide gel electrophoresis of TRCF-β and UvrB-β, respectively.

Fig. 34 is an alignment of the amino acid sequence of TRCF-β with that of UvrB-β.

Fig. 35 presents charts showing the SD spectra of TRCF-β and UvrB-β, respectively.

Fig. 36 presents charts showing the thermostabilities of TRCF-β and UvrB-β, respectively.

Fig. 37 presents charts showing the pH stabilities of TRCF-β and UvrB-β, respectively.

Fig. 38 shows the results of measurement of TRCF interactions using a BIAcore system.

Fig. 39 shows the results of measurement of the interaction between TRCF and UvrA.

Like reference symbols in the various drawings indicate like elements.

DETAILED DESCRIPTION

The invention provides novel DNA repair enzymes and nucleic acids encoding them. As described above, it is important to clone a large number of genes of highly stable DNA repair enzymes derived from highly thermophilic bacteria in order to elucidate DNA repair mechanisms and to obtain findings useful in various fields. The present invention has been achieved using genes of DNA repair enzymes derived from highly thermophilic bacteria belonging to the genus *Thermus*, in particular *Thermus thermophilus*, that are highly thermostable and suitable for three-dimensional structural analysis or

molecular function analysis. These enzyme proteins were produced in a large scale and subjected to analysis of substrate recognition mechanism to thereby complete the invention.

One exemplary DNA repair enzyme of the invention is a MutY enzyme, having a molecular weight approximately 31 kDa to 36 kDa, with a sequence as shown in SEQ ID NO: 2. MutY recognizes A:GO mismatches, A:G mismatches and G:GO mismatches, and removes inappropriate bases. See Example section below.

One exemplary DNA repair enzyme of the invention is a RecJ enzyme, having exonuclease activity that degrades single-stranded DNA only in the 5' to 3' direction. It has a molecular weight of approximately 50 kDa, with a sequence as shown in SEQ ID NO: 4. RecJ has specificity to single-stranded DNA, and a Km value of 6.2 μM. See Example section below.

One exemplary DNA repair enzyme of the invention is a RecF enzyme, having a molecular weight of approximately 37.8 kDa to 22 kDa, with a sequence as shown in SEQ ID NO: 8. RecF prevents replication at damaged sites. Briefly, when damage has occurred in DNA and the reaction of a replication complex stops at that site, a complex of RecF-RecO-RecR proteins binds to the DNA (see Example section below). The Km value is 31 μM at 37°C and 32 μM at 25°C.

One exemplary DNA repair enzyme of the invention is TRCF. TRCF interacts with UvrA and promotes the repair of damage-containing transcribed strands (see Example section below). Nucleotide excision repair mechanism in prokaryotes is also described below. Briefly, the complex UvrAB recognizes a damaged site and binds thereto. Damage in transcribed strands is recognized by TRCF and UvrA. TRCF has a molecular weight of approximately 37.8 kDa, and the theoretical molecular weight of TRCF-β region that is believed to be the binding site for UvrA is approximately 14.4 kDa. TRCF has a sequence as shown in SEQ ID NO: 6.

DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

The term "nucleic acid" as used herein refers to a deoxyribonucleotide (DNA) or ribonucleotide (RNA) in either single- or double-stranded form. The term encompasses nucleic acids containing known analogues of natural nucleotides. The term encompasses mixed oligonucleotides comprising an RNA portion bearing 2'-O-alkyl substituents conjugated to a DNA portion via a phosphodiester linkage, see, e.g., U.S. Patent No. 5,013,830. The term also encompasses nucleic-acid-like structures with synthetic backbones. DNA backbone analogues provided by the invention include phosphodiester, phosphorothioate, phosphorodithioate, methylphosphonate, phosphoramidate, alkyl phosphotriester, sulfamate, 3'-thioacetal, methylene(methylimino), 3'-N-carbamate, morpholino carbamate, and peptide nucleic acids (PNAs); see Oligonucleotides and Analogues, a Practical Approach, edited by F. Eckstein, IRL Press at Oxford University Press (1991); Antisense Strategies, Annals of the New York Academy of Sciences, Volume 600, Eds. Baserga and Denhardt (NYAS 1992); Milligan (1993) J. Med. Chem. 36:1923-1937; Antisense Research and Applications (1993, CRC Press). PNAs contain non-ionic backbones, such as N-(2-aminoethyl) glycine units. Phosphorothioate linkages are described, e.g., by U.S. Patent Nos. 6,031,092; 6,001,982; 5,684,148; see also, WO 97/03211; WO 96/39154; Mata (1997) Toxicol. Appl. Pharmacol. 144:189-197. Other synthetic backbones encompassed by the term include methyl-phosphonate linkages or alternating methylphosphonate and phosphodiester linkages (see, e.g., U.S. Patent No. 5,962,674; Strauss-Soukup (1997) Biochemistry 36:8692-8698), and benzylphosphonate linkages (see, e.g., U.S. Patent No. 5,532,226; Samstag (1996) Antisense Nucleic Acid Drug Dev 6:153-156). The term nucleic acid is used interchangeably with gene, DNA, RNA, cDNA, mRNA, oligonucleotide primer, probe and amplification product.

The terms "polypeptide," "protein," and "peptide" include compositions of the invention that also include "analogs," or "conservative variants" and "mimetics" or "peptidomimetics" with structures and activity that substantially correspond to the polypeptide from which the variant was derived, as discussed in detail, below.

The terms "array" or "microarray" or "biochip" or "chip" as used herein is an article of manufacture, a device, comprising a plurality of immobilized target elements, each target element comprising a "cluster" or "biosite" or defined area comprising a nucleic acid molecule or polypeptide of the invention immobilized to a solid surface, as discussed in further detail, below.

Generation and Genetic Engineering of Nucleic Acids

This invention provides novel nucleic acids encoding DNA repair enzymes of the invention, including antisense sequences, expression vectors, probes, PCR primers and the like. As the genes and vectors of the invention can be made and expressed *in vitro* or *in vivo*, the invention provides for a variety of means of making and expressing these genes and vectors. One of skill will recognize that desired phenotypes for altering and controlling nucleic acid expression can be obtained by modulating the expression or activity of the genes and nucleic acids (e.g., promoters, enhancers and the like) within the vectors of the invention. Any of the known methods described for increasing or decreasing expression or activity can be used for this invention. The invention can be practiced in conjunction with any method or protocol known in the art, which are well described in the scientific and patent literature.

The nucleic acid sequences of the invention and other nucleic acids used to practice this invention, whether RNA, cDNA, genomic DNA, vectors, viruses or hybrids thereof, may be isolated from a variety of sources, genetically engineered, amplified, and/or expressed recombinantly. Any recombinant expression system can be used, including, in addition to mammalian cells, e.g., bacterial, yeast, insect or plant systems.

Alternatively, these nucleic acids can be synthesized *in vitro* by well-known chemical synthesis techniques, as described in, e.g., Carruthers (1982) Cold Spring Harbor Symp. Quant. Biol. 47:411-418; Adams (1983) J. Am. Chem. Soc. 105:661; Belousov (1997) Nucleic Acids Res. 25:3440-3444; Frenkel (1995) Free Radic. Biol. Med. 19:373-380; Blommers (1994) Biochemistry 33:7886-7896; Narang (1979) Meth. Enzymol. 68:90; Brown (1979) Meth. Enzymol. 68:109; Beaucage (1981) Tetra. Lett. 22:1859; U.S. Patent No. 4,458,066. Double stranded DNA fragments may then be obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

Techniques for the manipulation of nucleic acids, such as, e.g., generating mutations in sequences, subcloning, labeling probes, sequencing, hybridization and the like are well described in the scientific and patent literature, see, e.g., Sambrook, ed., MOLECULAR CLONING: A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel, ed. John Wiley & Sons, Inc., New York (1997); LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY: HYBRIDIZATION WITH NUCLEIC ACID PROBES, Part I. Theory and Nucleic Acid Preparation, Tijssen, ed. Elsevier, N.Y. (1993).

Nucleic acids, vectors, capsids, polypeptides, and the like can be analyzed and quantified by any of a number of general means well known to those of skill in the art. These include, e.g., analytical biochemical methods such as NMR, spectrophotometry, radiography, electrophoresis, capillary electrophoresis, high performance liquid chromatography (HPLC), thin layer chromatography (TLC), and hyperdiffusion chromatography, various immunological methods, e.g. fluid or gel precipitin reactions, immunodiffusion, immuno-electrophoresis, radioimmunoassays (RIAs), enzyme-linked immunosorbent assays (ELISAs), immuno-fluorescent assays, Southern analysis, Northern analysis, dot-blot analysis, gel electrophoresis (e.g., SDS-PAGE), RT-PCR, quantitative PCR, other nucleic acid or target or signal amplification methods, radiolabeling, scintillation counting, and affinity chromatography.

In addition to "full length" DNA repair enzyme sequences (as determined by identity to the exemplary sequences of the invention, or, by functional criteria, e.g., based on a DNA repair activity, as described in detail in the examples, below), the invention also provides nucleic acid and polypeptides molecules that are only a portion of a "full length" sequence. For example, such a nucleic acid molecule can include a subsequence or fragment which can be used as a probe or primer or a fragment encoding a portion of a DNA repair enzyme domain, e.g., an immunogenic or biologically active portion of a DNA repair enzyme of the invention.

In another aspect, a nucleic acid of the invention includes a nucleotide sequence that includes part, or all, of the coding region and extends into either (or both) the 5' or 3' noncoding region, including both transcribed and non-transcribed sequences. Other embodiments include a fragment that includes a nucleotide sequence encoding an amino acid

fragment described herein. Nucleic acid fragments can encode a specific domain or site described herein or fragments thereof.

DNA repair enzyme probes and primers are provided. Typically a probe/primer is an isolated or purified oligonucleotide. The oligonucleotide typically includes a region of nucleotide sequence that hybridizes under stringent conditions (see below) to at least about 7, about 12, about 15, about 20, about 25, about 30, about 35, about 40, about 45, about 50, about 55, about 60, about 65, or about 75 consecutive nucleotides of a sense or antisense sequence of the exemplary sequences described herein. In one embodiment, the nucleic acid is a probe which is at least about 5 or about 10, and less than about 200 or less than 100 or less than 50 base pairs in length. In various embodiment, the probe or primer can be identical, or differ by 1, or less than about 5 or about 10 bases, from an exemplary sequence of the invention (while still capable of hybridizing under stringent conditions). If alignment is needed for this comparison the sequences can be aligned for maximum homology. "Looped" out sequences from deletions or insertions, or mismatches, are considered differences.

Amplification of Nucleic Acids

The invention provides oligonucleotide primers that can amplify DNA repair enzyme nucleic acids of the invention. The term "amplifying" and "amplification" as used herein incorporates its common usage and refers to the use of any suitable amplification methodology for generating or detecting recombinant or naturally expressed nucleic acid. For example, the invention provides methods and reagents (*e.g.*, specific degenerate oligonucleotide primer pairs) for amplifying (*e.g.*, by polymerase chain reaction, PCR) naturally expressed (e.g., genomic or mRNA) or recombinant (e.g., cDNA) nucleic acids of the invention *in vivo* or *in vitro*.

The nucleic acids of the invention can also be cloned or measured quantitatively using amplification techniques. Using the exemplary degenerate primer pair sequences of the invention (see below), the skilled artisan can select and design suitable oligonucleotide amplification primers. Amplification methods are also well known in the art, and include, *e.g.*, polymerase chain reaction, PCR (PCR PROTOCOLS, A GUIDE TO METHODS AND APPLICATIONS, ed. Innis, Academic Press, N.Y. (1990) and PCR STRATEGIES (1995), ed. Innis, Academic Press, Inc., N.Y., ligase chain reaction (LCR)

(see, e.g., Wu (1989) Genomics 4:560; Landegren (1988) Science 241:1077; Barringer (1990) Gene 89:117); transcription amplification (see, e.g., Kwoh (1989) Proc. Natl. Acad. Sci. USA 86:1173); and, self-sustained sequence replication (see, e.g., Guatelli (1990) Proc. Natl. Acad. Sci. USA 87:1874); Q Beta replicase amplification (see, e.g., Smith (1997) J. Clin. Microbiol. 35:1477-1491), automated Q-beta replicase amplification assay (see, e.g., Burg (1996) Mol. Cell. Probes 10:257-271) and other RNA polymerase mediated techniques (*e.g.*, NASBA, Cangene, Mississauga, Ontario); see also Berger (1987) Methods Enzymol. 152:307-316; Sambrook; Ausubel; U.S. Patent Nos. 4,683,195 and 4,683,202; Sooknanan (1995) Biotechnology 13:563-564.

Once amplified, the libraries can be cloned, if desired, into any of a variety of vectors using routine molecular biological methods; methods for cloning *in vitro* amplified nucleic acids are described, *e.g.*, U.S. Pat. No. 5,426,039. To facilitate cloning of amplified sequences, restriction enzyme sites can be "built into" the PCR primer pair. The primers can encode amino acid residues that are conservative substitutions (e.g., hydrophobic for hydrophobic residue) or functionally benign substitutions (e.g., retaining DNA repair activity).

Paradigms to design degenerate primer pairs are well known in the art. For example, a COnsensus-DEgenerate Hybrid Oligonucleotide Primer (CODEHOP) strategy computer program can be directly linked from the BlockMaker™ multiple sequence alignment site for hybrid primer prediction beginning with a set of related protein sequences. Means to synthesize oligonucleotide primer pairs are well known in the art. "Natural" base pairs or synthetic base pairs can be used. For example, use of artificial nucleobases offers a versatile approach to manipulate primer sequence and generate a more complex mixture of amplification products. Various families of artificial nucleobases are capable of assuming multiple hydrogen bonding orientations through internal bond rotations to provide a means for degenerate molecular recognition. Incorporation of these analogs into a single position of a PCR primer allows for generation of a complex library of amplification products. See, e.g., Hoops (1997) Nucleic Acids Res. 25:4866-4871. Nonpolar molecules can also be used to mimic the shape of natural DNA bases. A non-hydrogen-bonding shape mimic for adenine can replicate efficiently and selectively against a nonpolar shape mimic for thymine (see, e.g., Morales (1998) Nat. Struct. Biol. 5:950-954).

The invention provides sets of amplification primers capable of amplifying all or a portion of any DNA repair enzyme nucleic acid sequence of the invention, particularly, the exemplary sequence described herein. Thus, in one embodiment a set (pair) of primers is provided, e.g., primers suitable for use in a PCR, which can be used to amplify a selected region of a DNA repair enzyme sequence. In various embodiment, the primers can be at least about 5, about 10, or about 50 base pairs in length and can be less than about 100, or less than about 200, base pairs in length. The primers can be identical, or differ by one or more base residues from an exemplary sequence of the invention.

Generating Nucleic Acids from Cells

The invention provides method for generating nucleic acids that encode DNA repair enzymes by, e.g., amplification (e.g., PCR) of appropriate nucleic acid sequences using degenerate primer pairs, or traditional cloning using cDNA or genomic libraries, or, phage display libraries, or the like.

Genetic engineering of DNA Repair Enzyme-Encoding Sequences

The nucleic acid sequences of the invention can be operably linked to transcriptional or translational control elements, e.g., transcription and translation initiation sequences, promoters and enhancers, transcription and translation terminators, polyadenylation sequences, and other sequences useful for transcribing DNA into RNA. In construction of recombinant expression cassettes, vectors, transgenics, of the invention, a promoter fragment can be employed to direct expression of the desired nucleic acid in all tissues. Transcriptional or translational control elements can be isolated from natural sources, obtained from such sources as ATCC or GenBank libraries, or prepared by synthetic or recombinant methods.

The term "expression vector" refers to any recombinant expression system for the purpose of expressing a nucleic acid sequence of the invention *in vitro* or *in vivo*, constitutively or inducibly, in any cell, including prokaryotic, yeast, fungal, plant, insect or mammalian cell. The term includes linear or circular expression systems. The term includes expression systems that remain episomal or integrate into the host cell genome. The expression systems can have the ability to self-replicate or not, *i.e.*, drive only transient expression in a cell. The term includes recombinant expression "cassettes" which contain only the minimum elements needed for transcription of the recombinant nucleic acid.

The invention also provides fusion proteins comprising the polypeptides of the invention and heterologous domains, e.g., for protein detection, purification, or other applications. Detection and purification facilitating domains include, e.g., metal chelating peptides such as polyhistidine tracts or histidine- tryptophan modules or other domains that allow purification on immobilized metals; maltose binding protein; protein A domains that allow purification on immobilized immunoglobulin; or the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp, Seattle WA).

The inclusion of a cleavable linker sequences such as Factor Xa (see, e.g., Ottavi (1998) Biochimie 80:289-293), subtilisin protease recognition motif (see, e.g., Polyak (1997) Protein Eng. 10:615-619); enterokinase (Invitrogen, San Diego CA), and the like, can be useful to facilitate purification. For example, one construct can include a polypeptide-encoding nucleic acid sequence linked to six histidine residues followed by a thioredoxin, an enterokinase cleavage site (see, e.g., Williams (1995) Biochemistry 34:1787-1797), and an amino terminal translocation domain. The histidine residues facilitate detection and purification while the enterokinase cleavage site provides a means for purifying the desired protein(s) from the remainder of the fusion protein. Technology pertaining to vectors encoding fusion proteins and application of fusion proteins are well described in the scientific and patent literature, see e.g., Kroll (1993) DNA Cell. Biol., 12:441-53.

Cloning and construction of expression vectors

The invention provides expression vectors comprising the DNA repair enzyme nucleic acid sequences of the invention. These nucleic acids may be introduced into a genome or into the cytoplasm or a nucleus of a cell and expressed by a variety of conventional techniques, well described in the scientific and patent literature. See, e.g., Roberts (1987) Nature 328:731; Schneider (1995) Protein Expr. Purif. 6435:10; Sambrook, Tijssen or Ausubel. Product information from manufacturers of biological reagents and experimental equipment also provide information regarding known biological methods. The vectors can be isolated from natural sources, obtained from such sources as ATCC or GenBank libraries, or prepared by synthetic or recombinant methods.

The nucleic acids of the invention can be expressed in expression cassettes, vectors or viruses which are stably or transiently expressed in cells (e.g., episomal expression systems). Selection markers can be incorporated into expression cassettes and vectors to

confer a selectable phenotype on transformed cells and sequences. For example, selection markers can code for episomal maintenance and replication such that integration into the host genome is not required. For example, the marker may encode antibiotic resistance (e.g., chloramphenicol, kanamycin, G418, bleomycin, hygromycin) to permit selection of those cells transformed with the desired DNA sequences.

Inhibitory Sequences

The invention further provides for nucleic acids complementary to, *i.e.*, antisense sequences to, the DNA repair enzyme sequences of the invention. Antisense sequences are capable of inhibiting the transport, splicing or transcription of DNA repair enzyme-encoding genes. The inhibition can be effected through the targeting of genomic DNA or messenger RNA. The transcription or function of targeted nucleic acid can be inhibited, e.g, by hybridization and/or cleavage. One particularly useful set of inhibitors provided by the present invention includes oligonucleotides that are able to either bind DNA repair enzyme gene or message, in either case preventing or inhibiting the production or function of DNA repair enzymes. The association can be though sequence specific hybridization. Such inhibitory nucleic acid sequences can, e.g., be used to completely inhibit or depress the ability of DNA repair enzymes to repair DNA. Another useful class of inhibitors includes oligonucleotides that cause inactivation or cleavage of message. The oligonucleotide can have enzyme activity that causes such cleavage, such as ribozymes. The oligonucleotide can be chemically modified or conjugated to an enzyme or composition capable of cleaving the complementary nucleic acid. One may screen a pool of many different such oligonucleotides for those with the desired activity.

The invention provides for with antisense oligonucleotides capable of binding message that can inhibit DNA repair enzyme activity by targeting mRNA. Strategies for designing antisense oligonucleotides are well described in the scientific and patent literature, and the skilled artisan can design such oligonucleotides using the novel reagents of the invention. In some situations, naturally occurring nucleic acids used as antisense oligonucleotides may need to be relatively long (18 to 40 nucleotides) and present at high concentrations. A wide variety of synthetic, non-naturally occurring nucleotide and nucleic acid analogues are known which can address this potential problem. For example, peptide nucleic acids (PNAs) containing non-ionic backbones, such as N-(2-aminoethyl) glycine

units can be used. Antisense oligonucleotides having phosphorothioate linkages can also be used, as described in WO 97/03211; WO 96/39154; Mata (1997) *Toxicol Appl Pharmacol* 144:189-197; Antisense Therapeutics, ed. Agrawal (Humana Press, Totowa, N.J., 1996). Antisense oligonucleotides having synthetic DNA backbone analogues provided by the invention can also include phosphoro-dithioate, methylphosphonate, phosphoramidate, alkyl phosphotriester, sulfamate, 3'-thioacetal, methylene(methylimino), 3'-N-carbamate, and morpholino carbamate nucleic acids, as described herein.

Combinatorial chemistry methodology can be used to create vast numbers of oligonucleotides that can be rapidly screened for specific oligonucleotides that have appropriate binding affinities and specificities toward any target, such as the sense and antisense DNA repair enzyme sequences of the invention (for general background information, see, *e.g.*, Gold (1995) *J. of Biol. Chem.* 270:13581-13584). Combinatorial chemistry methodology can also be used to screen for agonist or antagonist ligands for DNA repair enzymes.

In yet another embodiment, the antisense nucleic acid molecule of the invention can be α-anomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β-units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids. Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330).

The invention provides for with ribozymes capable of binding DNA repair enzyme message which can inhibit DNA repair enzyme activity by targeting mRNA. Strategies for designing ribozymes and selecting the DNA repair enzyme-specific antisense sequence for targeting are well described in the scientific and patent literature, and the skilled artisan can design such ribozymes using the novel reagents of the invention. Ribozymes act by binding to a target RNA through the target RNA binding portion of a ribozyme which is held in close proximity to an enzymatic portion of the RNA that cleaves the target RNA. Thus, the ribozyme recognizes and binds a target RNA through complementary base-pairing, and once bound to the correct site, acts enzymatically to cleave and inactivate the target

RNA. Cleavage of a target RNA in such a manner will destroy its ability to direct synthesis of an encoded protein if the cleavage occurs in the coding sequence.

The enzymatic ribozyme RNA molecule can be formed in a hammerhead motif, but may also be formed in the motif of a hairpin, hepatitis delta virus, group I intron or RNaseP-like RNA (in association with an RNA guide sequence). Examples of such hammerhead motifs are described by Rossi (1992) *Aids Research and Human Retroviruses* 8:183; hairpin motifs by Hampel (1989) *Biochemistry* 28:4929, and Hampel (1990) *Nuc. Acids Res.* 18:299; the hepatitis delta virus motif by Perrotta (1992) *Biochemistry* 31:16; the RNaseP motif by Guerrier-Takada (1983) *Cell* 35:849; and the group I intron by Cech U.S. Pat. No. 4,987,071. The recitation of these specific motifs is not intended to be limiting; those skilled in the art will recognize that an enzymatic RNA molecule of this invention has a specific substrate binding site complementary to one or more of the target gene RNA regions, and has nucleotide sequence within or surrounding that substrate binding site which imparts an RNA cleaving activity to the molecule.

The inhibitory (e.g., antisense) nucleic acid molecules of the invention are typically administered to a subject (e.g., by direct injection at a tissue site), or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a DNA repair enzyme to thereby inhibit expression of the protein, e.g., by inhibiting transcription and/or translation. Alternatively, inhibitory nucleic acid molecules can be modified to target selected cells and then administered systemically. For systemic administration, inhibitory molecules can be conjugated with carriers that specifically bind to a receptor or an antigen expressed on a selected cell surface, e.g., by linking the inhibitory nucleic acid molecules to peptides or antibodies that bind to DNA repair enzymes or antigens. This linking can be direct or indirect, e.g., as by using liposomes. The inhibitory nucleic acid molecules can also be delivered to cells using the vectors such as viruses. To achieve sufficient intracellular concentrations of the inhibitory molecules, vector constructs in which the inhibitory nucleic acid molecule is placed under the control of a strong constitutive or inducible promoter, e.g., a pol II or a pol III promoter.

In other embodiments, a nucleic acid of the invention can also include other appended groups such as peptides (e.g., for targeting host cells *in vivo*), or agents facilitating transport across the cell membrane (see, e.g., Letsinger et al. (1989) *Proc. Natl. Acad. Sci.*

USA 86:6553-6556; Lemaitre et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W088/09810) or the blood-brain barrier (see, e.g., PCT Publication No. W089/10134). In addition, nucleic acids (e.g., oligonucleotides) can be modified with hybridization-triggered cleavage agents (See, e.g., Krol (1988) *Bio-Techniques* 6:958-976) or intercalating agents. (See, e.g., Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, (e.g., a peptide, hybridization triggered cross-linking agent, transport agent, or hybridization-triggered cleavage agent).

Transformed Bacterium, Transgenic and "Knockout" Cells and Organisms

The invention provides non-human transgenic (i.e., transformed) bacteria, animals and plants comprising the DNA repair enzyme nucleic acids of the invention or DNA repair enzyme "knockout" bacterial and animals generated using the nucleic acids of the invention. Such bacteria and animals are useful for studying the function and/or activity of DNA repair enzymes and for identifying and/or evaluating natural ligand, second messengers, modulators and other ligands of DNA repair enzyme activity. As used herein, a "transgenic animal" is a non-human animal, e.g., a mammal or a rodent, such as a rat or mouse, in which one or more of the cells of the animal includes a transgene (or is a "knockout"). Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, and the like. A transgene as used herein includes, e.g., exogenous DNA or a rearrangement, e.g., a deletion of endogenous chromosomal DNA, which preferably is integrated into or occurs in the genome of the cells of a transgenic bacteria or animal. A transgene can direct the expression of an encoded gene product in one cell (e.g., in a bacterium), or cells or tissues of a transgenic animal, other transgenes, e.g., a knockout, to reduce expression. Thus, a transgenic bacteria or animal can be one in which an endogenous DNA repair enzyme gene has been altered by, e.g., by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell, e.g., a bacterium or, an embryonic cell of an animal, prior to development of the animal.

DNA repair enzymes

The invention provides DNA repair enzymes, peptides, and fusion protein comprising these proteins, or subsequences thereof. An "isolated" or "purified" polypeptide or protein is substantially free of cellular material or other contaminating proteins from the

cell or tissue source from which the protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. In one embodiment, the language "substantially free" means preparation of DNA repair enzyme having less than about 30%, 20%, 10% and more preferably 5% (by dry weight), of non- DNA repair enzyme. When DNA repair enzymes or biologically active portions thereof are recombinantly produced, they can be prepared to be substantially free of culture medium, i.e., culture medium represents less than about 20%, or less than about 10%, or less than about 5% of the volume of the protein preparation. In alternative embodiments, the invention provides isolated or purified preparations of at least 0.01, 0.1, 1.0, and 10 milligrams in dry weight.

The invention provides DNA repair enzymes with non-essential amino acid residue substitutions. A "non-essential" amino acid residue is a residue that can be altered from the exemplary DNA repair enzyme sequences provided herein without abolishing or without substantially altering a binding or biological activity, whereas an "essential" amino acid residue results in such a change.

The invention provides DNA repair enzymes with conservative amino acid substitutions. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted nonessential amino acid residue in a DNA repair enzyme can be replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a DNA repair enzyme coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for DNA repair biological activity to identify mutants that retain activity. An alternative exemplary guideline uses the following six groups, each containing amino acids that are conservative substitutions for one another: 1) Alanine (A), Serine (S), Threonine (T); 2) Aspartic acid (D), Glutamic acid (E); 3)

Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); (see also, e.g., Creighton (1984) Proteins, W.H. Freeman and Company; Schulz and Schimer (1979) Principles of Protein Structure, Springer-Verlag). One of skill in the art will appreciate that the above-identified substitutions are not the only possible conservative substitutions. For example, for some purposes, one may regard all charged amino acids as conservative substitutions for each other whether they are positive or negative. In addition, individual substitutions, deletions or additions that alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence can also be considered "conservatively modified variations."

The invention also provides mimetic and peptidomimetic DNA repair enzymes. The terms "mimetic" and "peptidomimetic" refer to a synthetic chemical compound that has substantially the same structural and/or functional characteristics of enzymes, e.g., DNA repair. The mimetic can be either entirely composed of synthetic, non-natural analogues of amino acids, or, is a chimeric molecule of partly natural peptide amino acids and partly non-natural analogs of amino acids. The mimetic can also incorporate any amount of natural amino acid conservative substitutions as long as such substitutions also do not substantially alter the mimetic's structure and/or activity. As with polypeptides of the invention which are conservative variants, routine experimentation will determine whether a mimetic is within the scope of the invention, i.e., that its structure and/or function is not substantially altered. Polypeptide mimetic compositions can contain any combination of non-natural structural components, which are typically from three structural groups: a) residue linkage groups other than the natural amide bond ("peptide bond") linkages; b) non-natural residues in place of naturally occurring amino acid residues; or c) residues which induce secondary structural mimicry, i.e., to induce or stabilize a secondary structure, e.g., a beta turn, gamma turn, beta sheet, alpha helix conformation, and the like. A polypeptide can be characterized as a mimetic when all or some of its residues are joined by chemical means other than natural peptide bonds. Individual peptidomimetic residues can be joined by peptide bonds, other chemical bonds or coupling means, such as, e.g., glutaraldehyde, N-hydroxysuccinimide esters, bifunctional maleimides, N,N'-dicyclohexylcarbodiimide (DCC) or N,N'-diisopropylcarbodiimide (DIC). Linking groups that can be an alternative to the

traditional amide bond ("peptide bond") linkages include, e.g., ketomethylene (e.g., -C(=O)-CH_2- for -C(=O)-NH-), aminomethylene (CH_2-NH), ethylene, olefin (CH=CH), ether (CH_2-O), thioether (CH_2-S), tetrazole (CN_4-), thiazole, retroamide, thioamide, or ester (see, e.g., Spatola (1983) in Chemistry and Biochemistry of Amino Acids, Peptides and Proteins, Vol. 7, pp 267-357, "Peptide Backbone Modifications," Marcell Dekker, NY). A polypeptide can also be characterized as a mimetic by containing all or some non-natural residues in place of naturally occurring amino acid residues; non-natural residues are well described in the scientific and patent literature.

The invention provides polypeptides that are less than "full length" such that they only comprise a ligand domain for purposes of screening studies, directed mutagenesis, biological studies, as immunogens, for fusion proteins, and the like. As used herein, a "biologically active portion" of a DNA repair enzyme includes a fragment of a DNA repair enzyme that participates in a DNA repair activity. Biologically active portions of DNA repair enzymes include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequence of an exemplary DNA repair enzymes of the invention. These peptides can include less amino acids than "full length" DNA repair enzymes, and can exhibit at least one activity (e.g., DNA binding or biological activity or immunogenic property) of a "full length" DNA repair enzyme. Typically, biologically active portions comprise a complete domain or motif with at least one activity of the DNA repair enzyme, e.g., specific binding to a DNA base pair mismatch. A biologically active portion of a DNA repair enzyme can be a polypeptide that is, e.g., 10, 25, 50, 100, 200 or more amino acids in length. Biologically active portions of a DNA repair enzymes can be used as targets for developing agents which modulate a DNA repair enzyme mediated activity.

Fusion proteins of the invention can also include all or a part of a serum protein, e.g., an IgG constant region, or human serum albumin. The fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject *in vivo*. These fusion proteins can be used to affect the bioavailability of a DNA repair enzyme substrate or pharmaceutical composition. Fusion proteins as pharmaceutical compositions can be useful therapeutically for the treatment of disorders caused by, for example, (i) aberrant modification or mutation of a gene encoding a DNA repair enzyme; (ii)

mis-regulation of a DNA repair enzyme gene of the invention; and (iii) aberrant post-translational modification of a DNA repair enzyme.

Sequence homology determinations

The invention provides several subfamilies, or genuses, of nucleic acids and DNA repair enzymes (as set forth by the exemplary sequences of the invention, and as described in detail herein), members of which are determined to be within the scope of the invention by calculations of their homology, or sequence identity, to the exemplary sequences of the invention. To determine the percent identity of two amino acid sequences, or of two nucleic acid sequences (to determine if they are within the scope of the invention), the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In one embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, or at least 40%, or at least 50%, or at least 60%, or at least 70%, 80%, 90%, or 100% of the length of the reference sequence (e.g., when aligning a second sequence to exemplary DNA repair enzyme amino acid sequences. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In one embodiment, the percent identity between two amino acid sequences is determined using the algorithm described in Needleman (1970) *J. Mol. Biol.* (48):444-453, and variations thereof; this algorithm has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet

another embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6.

The percent identity between two amino acid or nucleotide sequences also can be determined using the algorithm of E. Meyers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences described herein can be used as a "query sequence" to perform a search against public databases to, e.g., to identify other DNA repair enzyme family members. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to DNA repair enzyme molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov.

Stringent Hybridization Methods

Nucleic acids with the scope of the invention can also be determined by their ability to hybridize to an exemplary nucleic acid of the invention by stringent hybridization. The phrase "stringent conditions" refers to hybridization or wash conditions under which a nucleic acid, e.g., a sample nucleic acid or a probe will primarily hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences in significant amounts. A positive signal (*e.g.*, identification of a nucleic acid of the invention) is about 10 times background hybridization. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in

Tijssen, Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH, and nucleic acid concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at Tm, 50% of the probes are occupied at equilibrium).

Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (*e.g.*, 10 to 50 nucleotides) and at least about 60°C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide.

Stringent hybridization conditions that can be used to identify nucleic acids within the scope of the invention can include hybridization in a buffer comprising 50% formamide, 5x SSC, and 1% SDS at 42°C, or hybridization in a buffer comprising 5x SSC and 1% SDS at 65°C, both with a wash of 0.2x SSC and 0.1% SDS at 65°C. Exemplary stringent hybridization conditions can also include a hybridization in a buffer of 40% formamide, 1 M NaCl, and 1% SDS at 37°C, and a wash in 1X SSC at 45°C. Alternatively, hybridization to filter-bound DNA in 0.5 M $NaHPO_4$, 7% sodium dodecyl sulfate (SDS), 1 mM EDTA at 65°C, and washing in 0.1X SSC/0.1% SDS at 68°C can be used to identify and isolate nucleic acids within the scope of the invention. Those of ordinary skill will readily recognize that alternative but comparable hybridization and wash conditions can be utilized to provide conditions of similar stringency.

However, the selection of a hybridization format is not critical, as is known in the art, it is the stringency of the wash conditions that set forth the conditions which determine whether a nucleic acid is within the scope of the invention. Wash conditions used to identify nucleic acids within the scope of the invention include, e.g.: a salt concentration of about 0.02 molar at pH 7 and a temperature of at least about 50°C or about 55°C to about 60°C; or, a salt concentration of about 0.15 M NaCl at 72°C for about 15 minutes; or, a salt

concentration of about 0.2X SSC at a temperature of at least about 50°C or about 55°C to about 60°C for about 15 to about 20 minutes; or, the hybridization complex is washed twice with a solution with a salt concentration of about 2X SSC containing 0.1% SDS at room temperature for 15 minutes and then washed twice by 0.1X SSC containing 0.1% SDS at 68°C for 15 minutes; or, equivalent conditions. Stringent conditions for washing can also be 0.2 X SSC/0.1% SDS at 42°C. In instances wherein the nucleic acid molecules are deoxyoligonucleotides ("oligos"), stringent conditions can include washing in 6X SSC/0.05% sodium pyrophosphate at 37°C (for 14_base oligos), 48°C (for 17-base oligos), 55°C (for 20-base oligos), and 60°C (for 23-base oligos). See Sambrook, ed., MOLECULAR CLONING: A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel, ed. John Wiley & Sons, Inc., New York (1997), or Tijssen (1993) supra, for detailed descriptions of equivalent hybridization and wash conditions and for reagents and buffers, e.g., SSC buffers and equivalent reagents and conditions.

Anti-DNA Repair Enzyme Antibodies

The invention also provides antibodies specifically reactive with the DNA repair enzymes of the invention. The term "antibody" as used herein refers to an immunoglobulin molecule or immunologically active portion thereof, i.e., an antigen-binding portion. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')$_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The antibody can be a polyclonal, monoclonal, recombinant, e.g., a chimeric or humanized, fully human, non-human, e.g., murine, or single chain antibody. In one embodiment the antibody has an effector function and can fix complement. The antibody can be coupled to a toxin or imaging agent.

A full-length DNA repair enzyme or, an antigenic peptide fragment thereof, can be used as an immunogen or can be used to identify anti- DNA repair enzyme antibodies made with other immunogens, e.g., cells, membrane preparations, and the like. In various embodiments, antigenic peptides of a DNA repair enzyme can include at least about 8, at least about 8, at least about 15, at least about 20, at least about 25, or at least about 30 amino acid residues of an exemplary sequence of the invention.

Subsequences or fragments of DNA repair enzyme can be used as immunogens or used to characterize the specificity of an antibody. In various embodiments, antibodies of the invention bind to hydrophilic regions of the protein, or, extracellular or, intracellular, or loop, or ligand or second messenger binding regions or motifs (and can also have agonist or antagonist activity). Antibodies reactive with, or specific for, any of these regions, or other regions or domains described herein are provided.

Exemplary epitopes encompassed by DNA repair enzyme antigenic peptides of the invention are regions located on the surface of the protein, e.g., hydrophilic regions, as well as regions with high antigenicity. For example, an Emini surface probability analysis of the protein sequence can be used to indicate the regions that have a particularly high probability of being localized to the surface or solvent (e.g., extracellular or intracellular fluids) of the protein and are thus likely to constitute surface residues useful for targeting antibody production.

Chimeric, humanized, or completely human antibodies are desirable for applications which include repeated administration, e.g., therapeutic treatment (and some diagnostic applications) of human patients. The anti- DNA repair enzyme antibody can be a single chain antibody. A single-chain antibody (scFV) may be engineered (see, for example, Colcher (1999). *Ann. N Y Acad. Sci.* 880:263-80; Reiter (1996) *Clin. Cancer Res.* 2:245-252. The single chain antibody can be dimerized or multimerized to generate multivalent antibodies having specificities for different epitopes of the same target DNA repair enzyme.

An antibody of the invention (e.g., monoclonal antibody or antiserum) can be used to isolate DNA repair enzymes by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, an anti-DNA repair enzyme antibody can be used to detect DNA repair enzymes (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the protein. Anti-DNA repair enzyme antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance (i.e., antibody labeling). Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish

peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Methods for identifying DNA repair enzyme agonists and antagonists

The invention provides methods (also referred to herein as "screening assays") for identifying modulators, i.e., candidate or test compounds or agents (e.g., proteins, peptides, peptidomimetics, peptoids, small molecules or other drugs) which bind to DNA repair enzymes, have a stimulatory or inhibitory effect on, e.g., DNA repair enzyme expression or activity, or have a stimulatory or inhibitory effect on, e.g., the expression or activity of a DNA repair enzyme. Compounds thus identified can be used to modulate the activity of target gene products (e.g., DNA repair enzyme genes) in a therapeutic protocol, to elaborate the biological function of the target gene product, or to identify compounds that disrupt normal target gene interactions. Exemplary protocols that can be used to measure DNA repair activity are well known in the art, see, e.g., the Examples, below.

The invention provides methods and compositions for determining whether a test compound specifically binds to a DNA repair enzyme *in vitro* or *in vivo*. The invention also provides methods and compositions for determining whether a test compound can effect the physiology of a cell expressing a DNA repair enzyme. Any aspect of cell physiology can be monitored to assess the effect of ligand binding to a DNA repair enzyme of the invention.

The invention also provides bacterium and non-human animals expressing one or more DNA repair enzyme sequences of the invention. Such expression can be used to determine whether a test compound specifically binds to a DNA repair enzyme *in vivo* by contacting a stably or transiently infected organism with a nucleic acid of the invention with a test compound and determining whether the cell or animal reacts to the test compound by specifically binding to the DNA repair enzyme.

The DNA repair enzymes of the invention can be expressed *in vivo* by delivery with an infecting agent, a vector, or a virus, e.g., adenovirus expression vector.

Bacterium and animals infected with the vectors of the invention are particularly useful for assays to identify and characterize ligands that can bind to (and act as antagonists or agonists) of subfamilies of DNA repair enzymes. Such vector-infected animals can be used for *in vivo* screening of putative ligands and their effect on, e.g., cell physiology, e.g., as with DNA repair.

Test compounds can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; peptoid libraries (libraries of molecules having the functionalities of peptides, but with a novel, non-peptide backbone which are resistant to enzymatic degradation but which nevertheless remain bioactive, see, e.g., Zuckermann (1994) *J. Med. Chem.* 37: 2678-85; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library and peptoid library approaches are limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (see, e.g., Lam (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:6909; Erb (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann (1994). *J. Med. Chem.* 37:2678; Cho (1993) *Science* 261:1303; Carrell (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; Gallop (1994) *J. Med. Chem.* 37:1233. Libraries of compounds may be presented in solution (see, e.g., Houghten (1992) *Biotechniques* 13:412-421), or on beads (see, e.g., Lam (1991) *Nature* 354:82-84), chips (see, e.g., Fodor (1993) *Nature* 364:555-556), bacteria (see, e.g., Ladner USP 5,223,409), spores (see, e.g., Ladner USP '409), plasmids (see, e.g., Cull (1992) *Proc. Natl. Acad. Sci. USA* 89:1865-1869) or on phage (see, e.g., Scott (1990) *Science* 249:386-390; Cwirla (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382; Felici (1991) *J. Mol. Biol.* 222:301-310).

In yet another embodiment, a cell-free assay is provided in which a DNA repair enzyme or biologically active portion thereof is contacted with a test compound and the ability of the test compound to bind to the DNA repair enzyme or biologically active portion thereof is evaluated. Biologically active portions of the DNA repair enzymes can be

used in assays of the present invention include fragments which participate in interactions with non-DNA repair enzymes, e.g., fragments with high surface probability scores.

Cell-free assays involve preparing a reaction mixture of the target gene protein and the test compound under conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex that can be removed and/or detected. The interaction between two molecules can also be detected, e.g., using fluorescence energy transfer (FET) (see, for example, Lakowicz *et al.*, U.S. Patent No. 5,631,169; Stavrianopoulos, et al., U.S. Patent No. 4,868,103). A fluorophore label on the first, 'donor' molecule is selected such that its emitted fluorescent energy will be absorbed by a fluorescent label on a second, 'acceptor' molecule, which in turn is able to fluoresce due to the absorbed energy. Alternately, the 'donor' protein molecule may simply utilize the natural fluorescent energy of tryptophan residues. Labels are chosen that emit different wavelengths of light, such that the 'acceptor' molecule label may be differentiated from that of the 'donor'. Since the efficiency of energy transfer between the labels is related to the distance separating the molecules, the spatial relationship between the molecules can be assessed. In a situation in which binding occurs between the molecules, the fluorescent emission of the 'acceptor' molecule label in the assay should be maximal. An FET binding event can be conveniently measured through standard fluorometric detection means well known in the art (e.g., using a fluorimeter).

Determining the ability of DNA repair enzymes to bind to a target molecule can be accomplished using real-time Biomolecular Interaction Analysis (BIA) (see, e.g., Sjolander, S. and Urbaniczky, C. (1991) *Anal. Chem.* 63:2338-2345 and Szabo et al. (1995) *Curr. Opin. Struct. Biol.* 5:699-705). "Surface plasmon resonance" or "BIA" detects biospecific interactions in real time, without labeling any of the interactants (e.g., BIAcore). Changes in the mass at the binding surface (indicative of a binding event) result in alterations of the refractive index of light near the surface (the optical phenomenon of surface plasmon resonance (SPR)), resulting in a detectable signal which can be used as an indication of real-time reactions between biological molecules.

The target gene product or the test substance can be anchored onto a solid phase. The target gene product/test compound complexes anchored on the solid phase can be detected at the end of the reaction. Preferably, the target gene product can be anchored onto

a solid surface, and the test compound, (which is not anchored), can be labeled, either directly or indirectly, with detectable labels discussed herein.

It may be desirable to immobilize either DNA repair enzymes, an anti-antibody or DNA repair enzyme target molecules to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to a DNA repair enzyme, or interaction of DNA repair enzymes with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase/ DNA repair enzyme fusion proteins or glutathione-S-transferase/target fusion proteins can be adsorbed onto glutathione Sepharose™ beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or DNA repair enzyme, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix immobilized in the case of beads, complex determined either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of DNA repair enzyme binding or activity determined using standard techniques.

Other techniques for immobilizing either DNA repair enzymes or a target molecule on matrices include using conjugation of biotin and streptavidin. Biotinylated DNA repair enzyme or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques known in the art (e.g., biotinylation kit, Pierce Chemicals, Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical).

Kits

The invention provides kits that contain DNA repair enzymes of the invention. The invention provides kits that contain oligonucleotide primer pairs and/or probes capable of amplifying and/or identifying nucleic acids of the invention. The kit can contain

instructional material teaching methodologies, e.g., means to repair DNA using the DNA repair enzymes of the invention.

In one embodiment, the kit can include a compound or agent capable of detecting a DNA repair enzyme of the invention or a corresponding mRNA in a biological sample. A standard can be included. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect DNA repair enzyme or nucleic acid.

For antibody-based kits, the kit can include: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide corresponding to a DNA repair enzyme of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

For oligonucleotide-based kits, the kit can include: (1) an oligonucleotide, e.g., a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide corresponding to a marker of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule corresponding to a marker of the invention. The kit can also includes a buffering agent, a preservative, or a protein stabilizing agent. The kit can also includes components necessary for detecting the detectable agent (e.g., an enzyme or a substrate).

The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit can be enclosed within an individual container and all of the various containers can be within a single package, along with instructions for interpreting the results of the assays performed using the kit.

EXAMPLES

The following examples are offered to illustrate, but not to limit the claimed invention.

Example 1: Isolation and Characterization of DNA Repair Enzyme Sequences

The following example describes the isolation and identification of the novel DNA Repair Enzyme sequences of the invention.

In nature, DNA is undergoing damage caused by endogenous factors, such as various types of active oxygen generated from energy metabolism or oxidation stress, and exogenous factors, such as ultraviolet light, ionizing radiation, or chemical substances. Further, mismatches that do not pair correctly with the template may be generated during DNA replication. For example, accurate DNA strands may not be synthesized in polymerase chain reaction (PCR) depending on the DNA polymerase used. The proteins of the invention are enzymes that repair these mismatches and bring about proper base pairs.

The DNA repair enzymes isolated in the present invention are the four enzymes of MutY, RecJ, RecF and TRCF.

(1) MutY

DNA in aerobic organisms is always being damaged by active oxygen generated from energy metabolism or stress. Guanine is susceptible to oxidation into 8-oxoguanine (GO), which not only pairs with cytosine but also mispairs with adenine during replication, giving rise to C:G to A:T transversion (Fig. 1). In order to prevent this mutation, MutY recognizes A:GO mismatches and removes adenine; recognizes G:GO mismatches and removes guanine; and also recognizes A:G mismatches and removes adenine.

Action: Steps of repairing are shown in Fig. 2 (Panels A-E). First, MutY removes the inappropriate base from the damaged site in DNA by its DNA glycosylase activity (Panel A). Then, MutY cuts the DNA strand on the 3' side of the base-removed site (AP site) by its AP lyase activity (Panel B). Finally, the gap is filled by the actions of esterase, DNA polymerase and DNA ligase. Thus, the repair is completed (Panel E).

Molecular Weight: The theoretical molecular weight of the MutY of the invention calculated from its amino acid sequence is 36 kDa; the molecular weight estimated from SDS-polyacrylamide gel electrophoresis is ~36 kDa (Fig. 3); and the molecular weight estimated from gel filtration (Superdex 200HR™, 50 mM Tris-HCl (pH 8), 0.5 M NaCl) is 31 kDa (Fig. 4).

Amino Acid Sequence: The sequence is shown in SEQ ID NO: 2. Comparison of this sequence with amino acid sequences of other microorganisms-derived MutY proteins reveals that the residue essential for N-glycosylase activity and residues constituting an iron-sulfur cluster are conserved (Fig. 5).

Substrate Specificity: MutY recognizes A:GO mismatches, A:G mismatches and G:GO mismatches, and removes inappropriate bases.

Absorption Spectrum: The results of measurement in solution containing 50 mM potassium phosphate (pH 7.5), 0.8 M KCl, 1 mM DTT, 1 mM EDTA and 10% glycerol revealed that MutY has a spectrum peculiar to an iron-sulfur cluster at around 410 nm (Fig. 8).

α-Helix Content: The results of CD spectrum analysis in a solution containing 50 mM Tris-HCl (pH 8.0), 0.1 M KCl, 1 mM DTE, 1 mM EDTA and 20% glycerol revealed that α-helix content is ~40% (Fig. 9).

Thermostability: The results of analysis of CD spectrum at varied temperatures in a solution containing 50 mM potassium phosphate (pH 7.5), 0.1 M KCl, 1 mM DTE, 1 mM EDTA and 20% glycerol revealed that MutY is stable at temperatures from 24oC to 80oC (especially, up to 75oC) under neutral conditions (pH 7.5) (Fig. 10).

(2) RecJ

RecJ is a DNA repair enzyme with both exonuclease activity specific to single-stranded DNA and deoxyribodiesterase activity, and is involved in both base excision repair system and mismatch repair system (Fig. 12). It is also known that RecJ carries out the initial process of homologous recombination in cooperation with RecQ and SSB (both of which are single-stranded DNA-binding proteins).

In base excision repair system, the function of RecJ is to cut the DNA strand on the 3' side of the nick generated by the actions of DNA glycosylase and AP endonuclease (Fig. 12, Left Panel).

In mismatch repair system, the function of RecJ is to degrade from the 5' to 3' direction the single-stranded DNA generated by the action of MutS, MutH, MutL or UvrD (Fig. 12, Central Panel).

In homologous recombination system, the function of RecJ is to degrade from the 5' to 3' direction the single-stranded DNA generated by the action of RecQ or SSB (Fig. 12, Right Panel).

RecJ homologues are found widely not only in prokaryotes but also in eukaryotes, such as yeast and Drosophila, and share characteristic motifs (Fig. 14).

Action: RecJ has exonuclease activity that degrades single-stranded DNA only in the 5' to 3' direction (Fig. 12).

Molecular Weight: The theoretical molecular weight of the RecJ of the invention calculated from its amino acid sequence is ~50 kDa, and the molecular weight estimated from SDS-polyacrylamide gel electrophoresis is ~50 kDa (Fig. 13).

Amino Acid Sequence: The sequence is shown in SEQ ID NO: 4.

Substrate Specificity: RecJ has specificity to single-stranded DNA, and the Km value is 6.2 μM (Figs. 17-23).

α-Helix Content: The results of CD spectrum analysis in a solution containing 50 mM K-Pi, 100 mM KCl, 0.1 mM DTE and 0.1 mM EDTA (pH 7.2) revealed that α-helix content is ~50% (Fig. 15).

Thermostability: The results of measurement of CD spectrum at varied temperatures using 1.6 μM RecJ in a solution containing 50 mM K-Pi, 100 mM KCl, 0.1 mM DTE and 0.1 mM EDTA (pH 7.2) revealed that the RecJ of the invention is stable up to 60oC (Fig. 16).

(3) RecF

From the results of genetic analyses so far made, it is known that RecF protein performs important functions in DNA recombinatorial repair, genetic recombination and DNA replication.

Action: In cooperation with RecO and RecR proteins, RecF prevents replication at damaged sites (Fig. 24). Briefly, when damage has occurred in DNA (Panel A) and the reaction of a replication complex stops at that site (Panel B), a complex of RecF-RecO-RecR proteins (RecFOR) binds to the DNA (Panel C). Then, replication re-starts (Panel D), and RecA causes pairing of homologous regions (Panel E), leading to strand exchange and DNA synthesis (Panel F). Subsequently, RuvA, RuvB and RuvC dissolve the Holliday structure formed by the pairing of homologous strands (a structure in which a homologous daughter strand is paired with each strand of a double-stranded DNA) to complete the repair (Panel G).

Molecular Weight: The theoretical molecular weight of the RecF of the invention calculated from its amino acid sequence is 37.8 kDa; the molecular weight estimated from gel filtration (Superdex 200HR™, 50 mM Tris-HCl, 2.0 M KCl (pH 7.5)) is

22 kDa (Fig. 26); and the molecular weight estimated from SDS-polyacrylamide gel electrophoresis is 37 kDa (Fig. 25).

Amino Acid Sequence: The sequence is shown in SEQ ID NO: 8. When this sequence is compared with amino acid sequences of other microorganisms-derived RecF proteins, high homology is observed partially (Fig. 27).

Substrate Specificity: The Km value is 31 μM at 37°C and 32 μM at 25°C.

α-Helix Content: The results of CD spectrum analysis in a solution containing 50 mM Tris-HCl and 100 mM KCl (pH7.2) revealed that α-helix content is ~40%.

Thermostability: The results of CD spectrum analysis revealed that RecF is stable up to ~50 °C at pH 7.5.

ATPase Activity: RecF, even alone, has ATPase activity (Fig. 29). This activity is increased when the substrate is single-stranded DNA, and decreased when the substrate is double-stranded DNA (Fig. 30).

(4) TRCF

Nucleotide excision repair performed by UvrA, UvrB and UvrC proteins is a mechanism which can recognize and remove DNA damage in the most wide range. Of these proteins, UvrA and UvrB form a complex, UvrAB, which specifically recognizes DNA damage. The results of three-dimensional structural analysis of UvrB revealed that a region that is believed to interact with UvrA forms one domain comprising β-sheet (UvrB-β) (Fig. 32) (Nakagawa et al., J. Biochem. 126, 986-990, 1999). TRCF (transcription-repair coupling factor) is a factor that interacts with UvrA and promotes the repair of damage-containing transcribed strands. TRCF has a region (TRCF-β homologous to the amino acid sequence of UvrB-β. This region is believed to be the binding site for UvrA.

Action: TRCF interacts with UvrA and promotes the repair of damage-containing transcribed strands (Fig. 31). Nucleotide excision repair mechanism in prokaryotes is as described below (Fig. 31). Briefly, the complex UvrAB recognizes a damaged site and binds thereto. Damage in transcribed strands is recognized by TRCF and UvrA. Then, the both ends of the damaged site are cut by the action of UvrC, and the site is removed. Subsequently, repair synthesis is completed by the actions of UvrD (helicase II), DNA polymerase I and DNA ligase.

Molecular Weight: The theoretical molecular weight of the TRCF of the invention calculated from its amino acid sequence is 37.8 kDa, and the theoretical molecular weight of TRCF-β region that is believed to be the binding site for UvrA is 14.4 kDa. The molecular weight of TRCF-β region estimated from SDS-polyacrylamide gel electrophoresis is 14.4 kDa (Fig. 33, Lower Panel).

Amino Acid Sequence: The sequence is shown in SEQ ID NO: 6. The amino acid sequences of the homologous regions between UvrB and TRCF (i.e., UvrB-β and TRCF-β) are highly conserved (Fig. 34).

CD Spectrum: The CD spectrum of TRCF-β measured in a buffer containing 50 mM Tris-HCl, 100 mM KCl (pH 7.9) resembles the CD spectrum of UvrB-β measured under the same conditions (Fig. 35).

Thermostability: The results of measurement of TRCF-β CD spectrum in a buffer containing 50 mM Tris-HCl, 100 mM KCl (pH 7.9) revealed that TRCF-β is stable at temperatures 20oC-75oC (Fig. 36).

pH Stability: The results of measurement of TRCF-β CD spectrum in various buffers containing 100 mM KCl revealed that TRCF-β is stable in a range from pH 4 to pH 9 at 25oC (Fig. 37).

The results of analysis of the interaction between TRCF and UvrA using a BIACORE sensor chip revealed that TRCF binds to UvrA. The dissociation constant of this binding is 0.5 μM in the presence of ATP and 1.3μM in the absence of ATP (Figs. 38 and 39).

Example 2: Isolation and Characterization of DNA Repair Gene Sequences

The following example describes the isolation (cloning) and identification of the novel DNA Repair Enzyme gene sequences of the invention.

The genes of the invention encode the above-described DNA repair enzymes. These genes can be obtained by the cloning technique described below. Hereinbelow, the cloning of the genes of the invention will be described specifically.

The genes of the invention can be isolated from the genomic DNA of *Thermus thermophilus*, a highly thermophilic bacterium.

Example 3: Preparation of DNA Repair Enzyme Genomic DNA

The following example describes the preparation of DNA Repair Enzyme genomic DNA sequences of the invention.

Genomic DNA may be prepared from cells of the above-mentioned bacterium by conventional methods. For example, cells are disrupted in a guanidine-containing buffer followed by phenol extraction to obtain crude DNA fraction. This fraction is subjected to cesium chloride gradient ultracentrifugation to obtain purified genomic DNA. The thus obtained genomic DNA is digested with an appropriate restriction enzyme (e.g., EcoRI, BamHI, or Sau3AI). For ligation of DNA fragments, T4 DNA ligase is used, for example.

DNA fragments treated with the above-mentioned restriction enzyme are ligated to a vector that has been digested with the same restriction enzyme used in the above treatment (e.g., EcoRI or BamHI) or a restriction enzyme that will generate a cohesive end complimentary to the digestion site generated by the enzyme used in the above treatment (e.g., BamHI against Sau3AI). It is also possible to construct a library from the resultant vector. Prior to the ligation, DNA fragments of interest may be amplified by PCR or the like. As a vector, a phage or plasmid capable of autonomous replication in a host organism is used. Specific examples of phage vector include EMBL3, M13 and λgt11. Specific examples of plasmid vector include pET systems (pET-3a, etc.), pBR systems (pBR322, etc.), pUC systems (pUC18, etc.) and pBluescript II (Stratagene). Further, various shuttle vectors may also be used in addition to those vectors capable of autonomous replication in two or more host organisms such as Escherichia coli or Bacillus subtilis. For the ligation of the DNA fragments and the vector fragments, a known DNA ligase (e.g., T4 DNA ligase) is used. The DNA fragments and vector fragments are ligated after annealing. The resultant vector is transferred into a host microorganism. DNA transfer into a host microorganism may be performed using any of conventional methods. For example, when the host is E. coli, such method as electroporation or the calcium phosphate method may be used. When a phage DNA is introduced into E. coli, an in vitro packaging method using a kit (Gigapack II™; Stratagene) may be used, for example.

Subsequently, host cells surviving in a medium containing antibiotics are screened by colony hybridization, etc. Plasmids are recovered from the selected host cells by

the alkali-SDS method or the like, to thereby obtain a genomic DNA fragment containing the gene of the invention.

The method of sequencing of the resultant DNA is not particularly limited. For example, a sequencing reaction may carried out using a PRISM™ sequencing kit containing a fluorescent dideoxyterminator (Perkin Elmer), followed by determination of the nucleotide sequence with an auto-sequencer from Applied Biosystems (e.g. Model ABI377).

In the present invention, MutY, RecJ, RecF and TRCF have been obtained as repair enzyme genes. SEQ ID NO: 1 shows the nucleotide sequence of the MutY gene of the invention, and SEQ ID NO: 2 shows the amino acid sequence encoded by this gene. SEQ ID NO: 3 shows the nucleotide sequence of the RecJ gene of the invention, and SEQ ID NO: 4 shows the amino acid sequence encoded by this gene. SEQ ID NO: 5 shows the nucleotide sequence of the RecF gene of the invention, and SEQ ID NO: 6 shows the amino acid sequence encoded by this gene. SEQ ID NO: 7 shows the nucleotide sequence of the TRCF gene of the invention, and SEQ ID NO: 8 shows the amino acid sequence encoded by this gene. It should be noted here that each of the above-mentioned amino acid sequences may have a mutation(s) such as deletion, substitution or addition of one or several amino acids, as long as a protein comprising that amino acid sequence retains DNA repair enzyme activity and is stable in a temperature range from 4°C to 100°C, up to 95°C; up to 90°C; up to 80°C, and up to 75°C.

For example, 1-10 amino acids, preferably 1-5 amino acids, may be deleted from the amino acid sequence as shown in SEQ ID NO: 2, 4, 6 or 8; 1-10 amino acids, preferably 1-5 amino acids, may be added to the amino acid sequence as shown in SEQ ID NO: 2, 4, 6 or 8; or 1-10 amino acids, preferably 1-5 amino acids, may be replaced with other amino acids in the amino acid sequence as shown in SEQ ID NO: 2, 4, 6 or 8. By creating mutants having such deletion, addition or substitution, it is possible to obtain proteins that are thermally more stable.

The term "DNA repair enzyme activity" used herein means activity that can recognize various types of damage caused in DNA and mismatch sites resulting from such damage, remove damaged sites or mismatch sites and fill the resultant gaps. Specific examples of target damage for repair include damage caused by active oxygen, damage

generated by UV irradiation, damage caused by chemical substances, and damage caused by PCR error.

The term "stability" used herein means that the structure of a protein as determined by CD spectrum analysis or the like is not changed up to 80 °C, preferably up to 75 °C, in a temperature range from 4 °C to 100 °C.

Also, the gene of the present invention may comprise a complementary strand to a DNA comprising the nucleotide sequence as shown in SEQ ID NO: 1, 3, 5 or 7.

Further, the gene of the present invention may comprise a DNA that can hybridize under stringent conditions either with the DNA repair enzyme gene or with a complementary strand thereto of the invention is included in the gene of the invention. Further, the gene of the present invention may comprise a DNA which hybridizes under stringent conditions with a probe prepared either from the above-described DNA of the invention (SEQ ID NO: 1, 3, 5 or 7) or from a complementary strand thereto, and which encodes a protein having DNA repair enzyme activity. The term "probe" used herein refers to a probe having a complementary sequence to the full-length sequence or a partial sequence consisting of at least 17 consecutive bases of the nucleotide sequence as shown in SEQ ID NO: 1, 3, 5 or 7. The term "stringent conditions" used herein refers to sodium concentrations between 15-300 mM, preferably 15-75 mM, and temperatures between 50-60 °C, preferably 55-60 °C.

Once the nucleotide sequence of the gene of the invention has been established, the gene of the invention can be obtained by chemical synthesis, by PCR using the cloned cDNA as a template, or by hybridization using a DNA fragment having the determined nucleotide sequence as a probe. Further, by using a technique such as site-specific mutagenesis, it is also possible to synthesize mutants of the gene of the invention that can express proteins with DNA repair enzyme activity.

Table 1

Substrate DNA	A:G	A:T				A:G O				A:C				A:G				G:G O			
MutY	-	+				+				+				+				+			
NaOH treatment	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-	+	+
Temperature (℃)	25	25	50	25	50	25	50	25	50	25	50	25	50	25	50	25	50	25	50	25	50
Lane No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21

These results demonstrate that MutY detected these mismatches and cut the substrate DNAs at the mismatch sites with its N-glycosylase activity and AP lyase activity.

Example 4: Preparation of *Thermus thermophilus* HB8-Derived *RecJ* Gene Product

Using genomic DNA from *Thermus thermophilus* HB8 as a template, a PCR reaction was carried out in the same manner as in Example 1, except that the following primers were used.

5' primer: 5'-ATCATATgAgAgACCgggTCCgCTggCgggT-3' (SEQ ID NO: 11)

3' primer: 5'-ATAgATCTTTACAggTCCACCgCCTggACCTC-3' (SEQ ID NO: 12)

A vector pET-19b that had been digested with *Nde*I and *Bam*HI and treated with a bacterial alkaline phosphatase for removal of its terminal phosphate group was ligated in a ligation reaction to the PCR product treated as described in Example 1 to thereby obtain a recombinant vector pET-19b-RecJ. Using this recombinant vector, *E. coli* BL21 (DE3) pLysE was transformed.

The nucleotide sequence of the gene encoding RecJ was determined in the same manner as in Example 1. As a result, the nucleotide sequence as shown in SEQ ID NO: 3 was obtained. The amino acid sequence encoded by this gene is shown in SEQ ID NO: 4.

The transformant prepared above was inoculated into 2 ml of LBamp medium and cultured at 37°C for 16 hrs. The resultant culture broth was added to 1 L of LBamp medium and cultured at 37°C for 3-4 hr. When cells reached the logarithmic phase, 50 μg/ml isopropyl-1-thio-β-D-galactoside (IPTG) was added thereto, followed by cultivation for 5-6 hr. The cells were harvested by centrifugation, washed with TE buffer and suspended in 20 ml of an adsorption buffer (20 mM Tris-HCl, 0.2 M NaCl, 5 mM imidazol and 1 mM 2-mercaptoethanol, pH 8.0), followed by sonication to disrupt cells. The resultant disrupted material was centrifuged at 10,000 g for 30 min to obtain a precipitate.

The thus obtained precipitate was dissolved in 6 M urea-containing adsorption buffer. Histidine-tagged RecJ protein in this solution was adsorbed onto chelating Sepharose. Briefly, the solution of the precipitate was added to chelating Sepharose that had been bound to Ni ions and washed sufficiently with 6 M urea-containing adsorption buffer. The resultant mixture was incubated at 4°C for 1 hr. Then, the Sepharose carrier was recovered by

centrifugation and washed sufficiently with the adsorption buffer. Subsequently, the Sepharose carrier was washed with adsorption buffers of gradually lowered urea concentrations (i.e., 4 M, 3 M, 2 M and 1 M) to thereby refold His-tagged RecJ protein. The RecJ protein was eluted with an elution buffer (20 mM Tris-HCl, 0.5 M NaCl, 500 mM imidazol, 1 mM 2-mercaptoethanol, pH 8.0). The purity of this His-tagged RecJ protein was confirmed by 12.5% SDS-polyacrylamide gel electrophoresis (Fig. 13). In Fig. 13, individual lanes are as follows:

M: molecular weight marker

Lane 1: total cell lysate

Lane 2: cell lysate (supernatant)

Lane 3: cell lysate (pellet)

Lane 4: chromatography fraction (Ni-NTA column)

Lane 5: refolding

Lane 6: anion exchange chromatography fraction (MonoQ column)

The bands indicated by an arrowhead represent His-tagged RecJ protein [$(His)_{10}$-RecJ].

Purified His-tagged RecJ protein was partially degraded with 100 units of thermolysin (Sigma) at 25°C for 6 hr to thereby obtain a soluble core domain with a molecular weight of 45 kDa.

Example 5: Physicochemical Properties of *Thermus thermophilus* HB8-Derived RecJ Protein

(1) CD Spectrum

CD spectrum was measured on 1.6 μM RecJ in a solution containing 50 mM potassium phosphate, 100 mM KCl, 0.1 mM DTE and 0.1 mM EDTA (pH 7.2). The results are shown in Fig. 15. From this Figure, it was found that the α-helix content of RecJ is ~50%.

(2) Thermostability Test

Thermostability was examined by analyzing the CD spectrum of the core domain obtained in Example 4 (1.6 μM) in a buffer containing 100 mM KCl, 0.1 mM dithiothreitol, 0.1 mM EDTA and 50 mM potassium phosphate (pH 7.5) while varying

temperatures. As a result, it was found that the core domain of RecJ protein is stable at temperatures from 15 °C to 60 °C (Fig. 16).

Example 6: Measurement of the Exonuclease Activity of *Thermus thermophilus* HB8-Derived RecJ Protein

The His-tagged RecJ protein obtained in Example 4 was degraded with thrombin to remove the tag. To a reaction solution (20 mM Tris-HCl, 10 mM $MgCl_2$, 100 mM KCl, 1 mM DTT, pH 7.5) containing 0.1 mM tag-removed RecJ protein, a single-stranded DNA of 49-mer (as shown below) whose 5' end had been labeled with a radioactive phosphate group was added as a substrate and reacted at 25 °C, 37 °C or 50 °C (Fig. 17).

Single-stranded DNA:

5'-ACTACTTggTACACTgACgCgAgCACgCAggAgCTCATTCCAgTgCGCA-3' (SEQ ID NO: 13)

The reaction products were analyzed by polyacrylamide gel electrophoresis. The results confirmed decrease of the substrate and increase of liberated, radioactive phosphate-labeled nucleotides with the passage of time. The results also indicated that RecJ protein has 5' to 3' exonuclease activity (Fig. 18).

Fig. 18 shows the 5' to 3' exonuclease activity of RecJ. Fig. 19 shows the dependency on RecJ concentration of the exonuclease activity. The exonuclease activity of RecJ increased depending on the RecJ concentration. Also, the activity increased further at a high temperature (50 °C).

Fig. 20 shows the results of examination of the effect of etheno-nucleotide upon RecJ exonuclease activity. Etheno-nucleotide is a fluorescently labeled nucleotide, which is characterized by emitting more intense fluorescence when it is liberated from DNA than when integrated in DNA. Thus, it is possible to know whether etheno-nucleotide has been liberated or not, i.e., whether DNA has been degraded or not, by measuring its fluorescence intensity. The RecJ exonuclease activity on etheno-nucleotide-labeled DNA and that on usual DNA were almost comparable. Thus, it was found that etheno-nucleotide-labeled DNA can be a substrate for RecJ.

Subsequently, a reaction solution containing 32μM etheno-nucleotide (εDNA), 0.4 μM RecJ, 20 mM Tris-HCl, 10 mM $MgCl_2$ and 100 mM KCl (pH 7.5) was incubated at

37°C, followed by detection of fluorescence with an excitation wavelength of 305 nm. The results are shown in Fig. 21 (lower panel titled "Fluorescent Spectrum"). Liberation of the etheno-nucleotide from DNA by the exonuclease activity of RecJ increased fluorescence intensity.

Further, a reaction solution containing 32μM etheno-nucleotide (εDNA), 0.4 μM RecJ, 20 mM Tris-HCl, 10 mM $MgCl_2$ and 100 mM KCl (pH 7.5) was incubated at 37 °C, followed by measurement of the time course of fluorescence intensity and the degree of fluorescence polarization with an excitation wavelength of 305 nm and a fluorescence wavelength of 410 nm.

The results are shown in Fig. 22. The upper panel shows the time course of fluorescence intensity, and the lower panel the time course of the degree of fluorescence polarization. When RecJ was reacted with etheno-nucleotide, the degree of fluorescence polarization that indicates the degree of freedom of fluorescent material increased. It is believed that this fact demonstrates the liberalization of the etheno-nucleotide from DNA.

Further, a reaction solution containing 0.1 μM RecJ, 20 mM Tris-HCl, 10 mM $MgCl_2$ and 100 mM KCl (pH 7.5) was incubated at 37 °C, followed by detection of fluorescence with an excitation wavelength of 305 nm and a fluorescence wavelength of 410 nm and measurement of the dependency of exonuclease activity upon DNA concentration.

The results are shown in Fig. 23. The results of calculation of kinetic parameters according to Michaelis-Menten equation were as follows: k_{cat} = 0.034/sec and K_m = 6.2 μM.

Example 7: Preparation of *Thermus thermophilus* HB8-Derived *RecF* Gene Product

Using genomic DNA from *Thermus thermophilus* HB8 as a template, a PCR reaction was carried out in the same manner as in Example 1, except that the following primers were used.

5' primer: 5'-ATATCATATgCgTCTTCTCCTCTTCCggCAACggAACT-3' (SEQ ID NO: 14)

3' primer: 5'-ATATAgATCTTTATTAggCgCCAgggCACAggACCACCCCT-3' (SEQ ID NO: 15)

A vector pET-15b that had been digested with *Nde*I and *Bam*HI and treated with a bacterial alkaline phosphatase for removal of its terminal phosphate group was ligated in a

ligation reaction to the PCR product treated as described in Example 1 to thereby obtain a recombinant vector pET-15b-RecF. Using this recombinant vector, *E. coli* BL21 (DE3) pLysE was transformed.

The nucleotide sequence of the gene encoding RecF was determined in the same manner as in Example 1. As a result, the nucleotide sequence as shown in SEQ ID NO: 5 was obtained. The amino acid sequence encoded by this gene is shown in SEQ ID NO: 6.

The transformant prepared above was inoculated into 2 ml of LBamp medium and cultured at 37 °C for 16 hrs. The resultant culture broth was added to 1 L of LBamp medium and cultured at 37 °C for 3-4 hr. When cells reached the logarithmic phase, 50 μg/ml isopropyl-1-thio-β-D-galactoside (IPTG) was added thereto, followed by cultivation for 5-6 hr. The cells were harvested by centrifugation, washed with TE buffer and suspended in 20 ml of an adsorption buffer (20 mM Tris-HCl, 0.2 M NaCl, 5 mM imidazol and 1 mM 2-mercaptoethanol, pH 8.0), followed by sonication to disrupt cells. The resultant disrupted material was centrifuged at 10,000 g for 30 min to obtain a supernatant.

His-tagged RecF protein in the resultant supernatant was adsorbed onto a chelating Sepharose column. Briefly, the supernatant was applied to a chelating Sepharose column that had been bound with Ni ions and equilibrated with the adsorption buffer. Then, the column was washed with the adsorption buffer. Subsequently, His-tagged RecF protein was eluted with an elution buffer (20 mM Tris-HCl, 0.2 M NaCl, 500 mM imidazol, 1 mM 2-mercaptoethanol, pH 8.0). The purity of this His-tagged RecF protein was confirmed by 12.5% SDS-polyacrylamide gel electrophoresis (Fig. 25).

In Fig. 25, individual lanes are as follows:

M: molecular weight marker

T: total cell lysate

S: cell lysate (supernatant)

His: histidine-tagged protein

HA: hydroxy apatite column chromatography fraction

Example 8: Physicochemical Properties of RecF

(1) CD Spectrum

CD spectrum was measured on 1.4 μM RecF in a solution containing 50 mM Tris-HCl and 100 mM KCl (pH 7.5). The results revealed that the α-helix content of RecF is ~40%.

(2) Thermostability Test

Thermostability was examined by analyzing CD spectrum in a solution containing 50 mM Tris-HCl and 100 mM KCl (pH 7.5) while varying temperatures.

The results revealed that RecF is stable up to 50 °C.

(3) Analysis of Binding Action

A reaction solution containing 5 μM RecF, 50 mM Tris-HCl, 100 mM KCl, 0.1 mM EDTA, and 5 mM 2-mercaptoethanol (pH 7.5) was incubated with εDNA at 25 °C, followed by analysis of fluorescence spectrum with an excitation wavelength of 310 nm. Since changes were observed in the spectrum of RecF in the presence of εDNA, it was found that RecF binds to DNA. The dissociation constant was 5.3 μM (Fig. 28).

(4) ATPase Activity

A reaction solution containing 1 μM RecF, 50 mM Tris-HCl (pH 7.5), 10 mM magnesium acetate, 100 mM KCl, 2 mM phosphoenolpyruvic acid, 0.3 mM NADH, 1 mM DTE, 25 U of pyruvate kinase and 25 U of lactate dehydrogenase was incubated, followed by measurement of ATPase activity. As a result, it was found that RecF, even alone, has ATPase activity and that this activity increases when the temperature is raised from 25 °C to 37 °C (Fig. 29).

Further, a reaction solution containing 1 μM RecF, 50 mM Tris-HCl (pH 7.5), 6 μM poly(dT) or 6 μM poly(dA) poly(dT), 10 mM magnesium acetate, 100 mM KCl, 2 mM phosphoenolpyruvic acid, 0.3 mM NADH, 1 mM DTE, 25 U of pyruvate kinase and 25 U of lactate dehydrogenase was incubated at 25 °C, followed by measurement of ATPase activity. The results revealed that ATPase activity increases in the presence of single-stranded DNA (poly(dT)) and decreases in the presence of double-stranded DNA (poly(dA)□poly(dT)) (Fig. 30).

Example 9: Preparation of *Thermus thermophilus* HB8-Derived *TRCF* (Transcription-Repair Coupling Factor) Gene Product

Using genomic DNA from *Thermus thermophilus* HB8 as a template, a PCR reaction was carried out in the same manner as in Example 1, except that the following primers were used.

5' primer: 5'-ATATCATATggAAATCgCgCTAgAgAggATCTACggCC-3' (SEQ ID NO: 16)

3' primer: 5'-ATATAgATCTTTATTAGAGGTCGGCGAAGAGGTAGAGCACC-3' (SEQ ID NO: 17)

A vector pET-15b that had been digested with *Nde*I and *Bam*HI and treated with a bacterial alkaline phosphatase for removal of its terminal phosphate group was ligated in a ligation reaction to the PCR product treated as described in Example 1 to thereby obtain a recombinant vector pET-15b-TRCF. Using this recombinant vector, *E. coli* BL21 (DE3) pLysE was transformed.

The nucleotide sequence of the gene encoding TRCF was determined in the same manner as in Example 1. As a result, the nucleotide sequence as shown in SEQ ID NO: 7 was obtained. The amino acid sequence encoded by this gene is shown in SEQ ID NO: 8.

The transformant prepared above was inoculated into 2 ml of LBamp medium and cultured at 37 °C for 16 hrs. The resultant culture broth was added to 1 L of LBamp medium and cultured at 37 °C for 3-4 hr. When cells reached the logarithmic phase, 50 μg/ml isopropyl-1-thio-β-D-galactoside (IPTG) was added thereto, followed by cultivation for 5-6 hr. The cells were harvested by centrifugation, washed with TE buffer and suspended in 20 ml of an adsorption buffer (20 mM Tris-HCl, 0.2 M NaCl, 5 mM imidazol and 1 mM 2-mercaptoethanol, pH 8.0), followed by sonication to disrupt cells. The resultant disrupted material was centrifuged at 10,000 g for 30 min to obtain a supernatant.

His-tagged TRCF protein in the resultant supernatant was adsorbed onto a chelating Sepharose column. Briefly, the supernatant was applied to a chelating Sepharose column that had been bound with Ni ions and equilibrated with the adsorption buffer. Then, the column was washed with the adsorption buffer. Subsequently, His-tagged TRCF protein was eluted with an elution buffer (20 mM Tris-HCl, 0.2 M NaCl, 500 mM imidazol, 1 mM 2-mercaptoethanol, pH 8.0). The purity of this His-tagged TRCF protein was confirmed by 12.5% SDS-polyacrylamide gel electrophoresis (Fig. 33). In Fig. 33, the upper panel shows

the results of purification of UvrB-β, and the lower panel shows the results of purification of TRCF-β. The lanes in the upper panel are as follows: M: molecular weight marker; 1: total cell lysate; 2: cell lysate (supernatant from centrifugation); 3: nickel column chromatography fraction; 4: butyl column chromatography fraction. The lanes in the lower panel are as follows: M: molecular weight marker; 1: total cell lysate; 2: nickel column & butyl column chromatography fraction.

Example 10: Physicochemical Properties of TRCF

(1) CD Spectrum

CD spectrum was measured on UvrB-β and TRCF-β at 25 °C in a solution containing 50 mM Tris-HCl and 100 mM KCl (pH 7.9). The results are shown in Fig. 35. It was found that UvrB-β and TRCF-β have similar three-dimensional structures.

(2) Thermostability Test

Thermostability was examined by analyzing the CD spectra of UvrB-β and TRCF-β in a solution containing 50 mM Tris-HCl and 100 mM KCl (pH 7.9) while varying temperatures. The results revealed that both UvrB-β and TRCF-β are stable at temperatures from 20 °C to 75 °C at pH 7.9 (Fig. 36).

(3) pH Stability

The CD spectra of UvrB-β and TRCF-β were measured in various buffers containing 100 mM KCl and having different pH values.

The results revealed that TRCF-β is stable at pH 4 to 9 at 25 °C (Fig. 37).

(4) Analysis of Binding Action

$NiCl_2$ was injected to a sensor chip NTA. Then, β domain was injected thereto and immobilized. Since it is known that UvrA and UvrB interact with each other only in the presence of ATP, the interaction between each β domain and UvrA was measured both in the presence of ATP and in the absence of ATP (Fig. 38).

As a result, it was found that the dissociation constant (K_d) is 0.5 μM in the presence of ATP and 1.3 μM in the absence of ATP (Fig. 39).

EFFECT OF THE INVENTION

According to the present invention, DNA repair enzymes and genes encoding the same are provided. The enzymes of the invention have DNA repair activity and are

excellent in thermostability. Therefore, they are useful as reagents for researches in molecular biology and other fields or as reagents for preventing or repairing errors in various DNA synthesis reactions.

SEQUENCE LISTING FREE TEXT

SEQ ID NO: 9: synthetic DNA

SEQ ID NO: 10: synthetic DNA

SEQ ID NO: 11: synthetic DNA

SEQ ID NO: 12: synthetic DNA

SEQ ID NO: 13: synthetic DNA

SEQ ID NO: 14: synthetic DNA

SEQ ID NO: 15: synthetic DNA

SEQ ID NO: 16: synthetic DNA

SEQ ID NO: 17: synthetic DNA